An approximation to the twin prime conjecture and the parity phenomenon

János Pintz*

1 Introduction

The celebrated theorem of Chen [Che1], [Che2] proved nearly 50 years ago asserts that there are infinitely many primes p for which $p+2$ is either a prime or has exactly two prime factors. In view of this strong approximation of the twin prime conjecture seems to be a surprise that it is not known whether there are infinitely many primes p such that $p+2$ has an odd number of prime factors. The reason for it is, as described by Hildebrand ([Hil]) "the so-called *parity barrier*, a heuristic principle according to which sieve methods cannot differentiate between integers with an even and odd number of prime factors." Iwaniec [Iwa] writes similarly about the *parity phenomenon*: "The parity phenomenon is best explained in the context of Bombieri's asymptotic sieve [Bom]. This says that within the classical conditions for the sieve one cannot sift but all numbers having the same parity of the number of prime divisors. Never mind producing primes; we cannot even produce numbers having either one, three, five or seven prime divisors. However, under the best circumstances we can obtain numbers having either 2006 or 2007 prime divisors. Similarly we can obtain numbers having either one or two prime divisors, but we are not able to determine which of these numbers are there, probably both."

In the present work we prove a weaker version of the problem that $\lambda(p+2)=-1$ for infinitely many primes p, where $\lambda(n)$ is Liouville's function:

$$\lambda(n)=(-1)^{\Omega(n)}, \quad \Omega(n)=\sum_{p|n} 1, \tag{1.1}$$

and as always, p, p_i, p' denote primes, $\mathcal{P}=\{p_n\}_{n=1}^{\infty}$ the set of all primes.

*Supported by OTKA Grants K72731, K67676 and ERC-AdG.228005.

Theorem 1. *There exists an even number d such that $0 < |d| \le 16$ and $\lambda(p+d) = -1$ for infinitely many primes p.*

The method proves actually somewhat more.

Theorem 1'. *At least one of the following two assertions is true:*

(i) *There exists an even d with $|d| = 2, 4$ or 8, such that $\lambda(p+d) = -1$ for infinitely many primes p;*

(ii) $\liminf_{n\to\infty}(p_{n+1} - p_n) \le 16$.

Remark. Theorem 1' remains true if $\lambda(p+d) = -1$ is replaced by $\lambda(p+d) = 1$.

An alternative version of Theorem 1 would be

Theorem 1". *There exists a positive even $d \le 18$ such that $\lambda(p+d) = -1$ for infinitely many primes p.*

Remark. The condition $2 \mid d$, $0 < d \le 18$ can be replaced here by $0 < d \le 17$.

With a refinement of the original method we can prove Theorem 1 with the additional requirement that $p+d$ should be an almost prime. Let $P^-(n)$ denote the smallest prime factor of n.

Theorem 2. *There exist absolute constants c and C, an* odd *$b \le C$ and a d with $0 < |d| \le 30$ such that $P^-(p+d) > p^c$, $\Omega(p+d) = b$ for infinitely many primes p.*

Similarly to Theorem 1' the modified method yields the following result:

Theorem 2'. *At least one of the following two assertions is true:*

(i) *There exist constants C, c, an* odd *$b \le C$ and an even $d \ne 0$ such that $|d| \le 18$ and $P^-(p+d) > p^c$, $\Omega(p+d) = b$ for infinitely many primes p;*

(ii) $\liminf_{n\to\infty}(p_{n+1} - p_n) \le 30$.

All the above results are based on the method [GPY] yielding

$$\liminf_{n\to\infty}(p_{n+1} - p_n)/\log p_n = 0, \tag{1.2}$$

and similarly to the above work we will investigate *admissible* k-element sets

$$\mathcal{H} = \{h_i\}_{i=1}^k, \quad 0 \le h_1 < h_2 < \dots < h_k, \quad h_i \in \mathbb{Z}, \tag{1.3}$$

where we call a set $\mathcal{H}$ admissible if it does not occupy all residue classes modulo any prime. In fact all the above mentioned results are simple consequences of more general ones referring to admissible sets.

In such a way, Theorems 1–1" are corollaries of the following result $(n + \mathcal{H} = \{n + h_i\}_{i=1}^k)$.

Theorem 3. *Let $\mathcal{H}$ be an admissible 6-tuple, $r \neq 0$, $r \notin \mathcal{H}$ any fixed integer. Then we have infinitely many integers n such that either*

(i) *$n + \mathcal{H}$ contains at least two primes, or*

(ii) *$n + h_i$ is prime for some $h_i \in \mathcal{H}$ and $\lambda(n + r) = -1$.*

It is easy to see that the set $\mathcal{H}_0 = \{0, 4, 6, 10, 12, 16\}$ is admissible. Choosing $r = 8$ we obtain Theorems 1 and 1', while the value $r = 18$ yields Theorem 1", $r = 17$ the result of the remark following Theorem 1".

Similarly to the above, Theorem 2' follows from

Theorem 4. *Let $\mathcal{H} = \{h_i\}_{i=1}^{9}$ be an admissible 9-tuple, $\mathcal{H}' = \mathcal{H} \setminus \{h_j\}$ with some fixed $j \in [1, 9]$. There exist absolute constants C, c, an* odd *integer $b \leq C$, such that we have infinitely many integers n with $P^-(n + h_i) > n^c$, $\Omega(n + h_i) \leq C$ for $1 \leq i \leq 9$, $\Omega(n + h_j) = b$ and a $\nu \in [1, 9]$, $\nu \neq j$ with $n + h_\nu \in \mathcal{P}$.*

It is easy to see that $\mathcal{H} = \mathcal{H}_1 = \{0, 2, 6, 8, 12, 18, 20, 26, 30\}$ is an admissible 9-tuple. This immediately gives Theorem 2, while the choice $h_j = 12$ (or $h_j = 18$) yields Theorem 2'.

2 Conditional theorems

A crucial ingredient of the proof is the celebrated Bombieri–Vinogradov theorem, similarly to the proof of (1.2). The number ϑ is called an admissible level of distribution of primes if for any $\varepsilon > 0$, $A > 0$

$$\sum_{q \leq N^{\vartheta - \varepsilon}} \max_{\substack{a \\ (a,q)=1}} \left| \sum_{\substack{p \equiv a \pmod q \\ p \leq N}} \log p - \frac{N}{\varphi(q)} \right| \ll_{\varepsilon, A} \frac{N}{\log^A N}. \tag{2.1}$$

The Bombieri–Vinogradov theorem asserts that $\vartheta = 1/2$ is an admissible level, while the Elliott–Halberstam conjecture states that $\vartheta = 1$ is admissible too. If ϑ is larger we can get closer to the original conjecture stating $\lambda(p + 2) = -1$ infinitely often. For $\vartheta > 0.729$ we can prove (see Theorem 7) the existence of infinitely many pairs n_1, n_2 with $|n_1 - n_2| \leq 2$, $n_1 \in \mathcal{P}$, $\lambda(n_2) = -1$. However, even assuming the Elliott–Halberstam conjecture we cannot prove the existence of infinitely many primes p with $\lambda(p + d) = -1$ for even a single a priori given d.

The possible conditional improvements over Theorems 1–4 depend on our knowledge of ϑ, the level of distribution of primes. However, we need

also an assumption about the λ-function, analogous to (2.1), namely we suppose the existence of a ϑ for which besides (2.1) also

$$\sum_{q\le N^{\vartheta-\varepsilon}} \max_{y\le N} \max_{a} \left| \sum_{\substack{n\equiv a \pmod q\\ n\le y}} \lambda(n) \right| \ll_{\varepsilon,A} \frac{N}{(\log N)^A} \tag{2.2}$$

holds. We will show analogues of Theorems 3–4 for the conditional case $\vartheta > 1/2$.

Theorem 5. *Let $\mathcal{H}$ be an admissible k-tuple, $k = C_1(\vartheta)$, $r \neq 0$, $r \notin \mathcal{H}$ any fixed integer. Then we have infinitely many integers n such that either*

(i) *$n + \mathcal{H}$ contains at least two primes, or*

(ii) *$n + h_i$ is prime for some $h_i \in \mathcal{H}$ and $\lambda(n + r) = -1$.*

The above holds with $C_1(0.729) = 2$, $C_1(0.616) = 3$, $C_1(0.554) = 4$ and $C_1(0.515) = 5$.

Theorem 6. *Let $\mathcal{H} = \{h_i\}_{i=1}^k$ be an admissible k-tuple, $k = C_2(\vartheta)$, $\mathcal{H}' = \mathcal{H}\backslash\{h_j\}$ for some fixed $j \in [1, C_2(\vartheta)]$. There exist absolute constants C, c, an* odd *$b \le C$ such that we have infinitely many integers n with $P^-(n+h_i) > n^c$, $\Omega(n + h_i) \le C$ for $1 \le i \le C_2(\vartheta)$, $\Omega(n + h_j) = b$ and a $\nu \in [1, C_2(\vartheta)]$ such that $\nu \neq j$, $n + h_\nu \in \mathcal{P}$. We can choose here $C_2(0.924) = 3$, $C_2(0.739) = 4$, $C_2(0.643) = 5$, $C_2(0.584) = 6$, $C_2(0.544) = 7$, $C_2(0.516) = 8$.*

The consequences of Theorem 5 (analogously to Theorems 1–1") are the following.

Theorem 7. *There exists an integer d, such that $0 < |d| \le C_3(\vartheta)$ and $\lambda(p+d) = -1$ for infinitely many primes p. We can choose here $C_3(0.729) = 2$, $C_3(0.616) = 6$, $C_3(0.554) = 8$, $C_3(0.515) = 12$.*

Remark. Apart from the first case $\vartheta = 0.729$ we can assume r to be even.

Theorem 7'. *At least one of the following two assertions is true:*

(i) *There exists a $|d| \le C_4(\vartheta)$ such that $\lambda(p+d) = -1$ for infinitely many primes p, where $C_4(0.729) = 1$, $C_4(0.616) = 3$, $C_4(0.554) = 4$, $C_4(0.515) = 7$.*

(ii) *$\liminf\limits_{n\to\infty}(p_{n+1} - p_n) \le C_3(\vartheta)$, with $C_3(\vartheta)$ as in Theorem 7.*

Remark. Theorem 7' remains true if $\lambda(p+d) = -1$ is replaced by $\lambda(p+d) = 1$ in (i).

Theorem 7". *There exists a positive even $d \le C_3(\vartheta)+2$ such that $\lambda(p+d) = -1$ infinitely often.*

Remark. The condition $2 \mid d$, $0 < d \le C_3(\vartheta) + 2$ can be replaced here by $0 < d \le C_3(\vartheta) + 1$.

The corollaries of Theorem 6 (analogously to Theorems 2–2') are the following.

Theorem 8. *There exist absolute constants C, c, an* odd $b \le C$ *and an even number d such that $0 < |d| \le C_5(\vartheta)$ and $P^-(p+d) > p^c$, $\Omega(p+d) = b$ for infinitely many primes p. We can choose here $C_5(0.924) = 6$, $C_5(0.739) = 8$, $C_5(0.643) = 12$, $C_5(0.584) = 16$, $C_5(0.544) = 20$, $C_5(0.516) = 26$.*

Theorem 8'. *At least one of the following assertions are true:*

(i) *There exist absolute constants C, c, an* odd $b \le C$ *and an even d such that $0 < |d| \le C_6(\vartheta)$ and $P^-(p+d) > p^c$, $\Omega(p+d) = b$ for infinitely many primes p. We can choose here $C_6(0.924) = 4$, $C_6(0.739) = 6$, $C_6(0.643) = 6$, $C_6(0.584) = 10$, $C_6(0.544) = 12$, $C_6(0.516) = 14$.*

(ii) $\liminf(p_{n+1} - p_n) \le C_5(\vartheta)$ *with the $C_5(\vartheta)$ given in Theorem 8.*

Remark. We emphasize here that under the strongest assumptions on ϑ we obtained the following assertions:

A) If $\vartheta > 0.729$, then we have a d with $0 < |d| \le 2$ such that $\lambda(p+d) = -1$ for infinitely many primes p.

B) If $\vartheta > 0.729$, then either the twin prime conjecture is true or $\lambda(p+d) = -1$, $|d| = 1$ holds for infinitely many primes p.

C) If $\vartheta > 0.924$, then there exist absolute constants C, c, an *odd* $b < C$ and a d with $0 < |d| \le 6$ such that $P^-(p+d) > p^c$, $\Omega(p+d) = b$ for infinitely many primes p.

D) If $\vartheta > 0.924$, then either $\liminf_{n\to\infty}(p_{n+1} - p_n) \le 6$ or there exists a C and an *odd* $b < C$ such that $\Omega(p+d) = b$ holds with $d = 2$ or $d = -4$ for infinitely many primes p (or alternatively one can choose $d = 4$ or $d = -2$).

In order to see that Theorems 5 and 6 imply the later results we have only to note that for $2 \le k \le 8$ we have the following admissible sets:

$$\begin{aligned}
\mathcal{H}_2 &= \{0, 2\},\\
\mathcal{H}_3 &= \{0, 2, 6\} \quad \text{or} \quad \{0, 4, 6\},\\
\mathcal{H}_4 &= \{0, 2, 6, 8\},\\
\mathcal{H}_5 &= \{0, 4, 6, 10, 12\},\\
\mathcal{H}_6 &= \{0, 4, 6, 10, 12, 16\},\\
\mathcal{H}_7 &= \{0, 2, 6, 8, 12, 18, 20\},\\
\mathcal{H}_8 &= \{0, 2, 6, 8, 12, 18, 20, 26\}.
\end{aligned}$$

Remark. One can show that for $k \leq 8$ these are the sets with minimal diameter, that is with minimal value of $h_k - h_1$ in (1.3).

In order to conclude Theorem 7 from Theorem 5 we can choose any $r \notin \mathcal{H}_k$ with $0 < r < h_k$. For Theorem 7' we choose r so that $|r - h_k/2|$ should be minimal, for Theorem 7" $r = h_k + 2$ (or $h_k + 1$ for the result in the remark after Theorem 7"). On the other hand, when deducing Theorem 8 from Theorem 6 we choose $r = h_j$ as any element of $\mathcal{H}_k$. To obtain Theorem 8' we choose it so that $|r - h_k/2|$ should be as small as possible.

Finally, if we would like to approximate the generalized twin prime problem ($p, p + 2d$ are both primes infinitely often for any integer $d > 0$), then we might consider the following two admissible sets for any $m \in \mathbb{Z}$:

$$\mathcal{H}_2' = \{0, 2d\}, \quad \mathcal{H}_3'' = \{0, 6m, 12m\}, \tag{2.3}$$

which yield the following corollaries to Theorems 5 and 6.

Corollary 1. *Assume $\vartheta > 0.729$. For any non-zero even d we have either*

(i) *infinitely many prime pairs $\{p, p + 2d\}$ or*

(ii) *infinitely many pairs $n_1, n_2 \in \mathbb{Z}$ with n_1 being prime, $\lambda(n_2) = -1$, $|n_1 - n_2| = d$.*

Proof. Choose $r = d$ in Theorem 5. □

Corollary 2. *Assume $\vartheta > 0.924$. For any $m \neq 0$ we have absolute constants C, c, an* odd *$b \leq C$ such that we have either*

(i) *infinitely many prime pairs $\{p, p + 12m\}$ or*

(ii) *infinitely many pairs $n_1, n_2 \in \mathbb{Z}$ with n_1 being prime, $P^-(n_2) > n_2^c$, $\Omega(n_2) = b$, $|n_1 - n_2| = 6m$.*

Proof. Choose $r = h_2 = 6m$ in Theorem 6. □

3 Proofs of Theorems 3 and 5

The idea of the proof is – analogously to [GPY] – to weigh the natural numbers with a weight inspired by Selberg's sieve (cf. (2.13) of [GPY])

$$\Lambda_R(n; \mathcal{H}, l) := \frac{1}{(k+l)!} \sum_{\substack{d \mid P_{\mathcal{H}}(n) \\ d \leq R}} \mu(d) \left(\log \frac{R}{d}\right)^{k+l}, \quad P_{\mathcal{H}}(n) = \prod_{i=1}^{k} (n + h_i), \tag{3.1}$$

more precisely by the square of a linear combination of these weights

$$a_n := a_n(\mathcal{H}; u) := \left(\Lambda_R(n; \mathcal{H}, 0) + \frac{u(k+1)}{\log R}\Lambda_R(n; \mathcal{H}, 1)\right)^2, \tag{3.2}$$

where u is a real parameter to be optimized according to the concrete problem. We will choose N as a large integer tending to infinity and let n run in the interval $[N, 2N)$ which we abbreviate by $n \sim N$. We will put

$$\chi_{\mathcal{P}}(n) := \begin{cases} 1 & \text{if } \ n \in \mathcal{P} \\ 0 & \text{if } \ n \notin \mathcal{P} \end{cases}, \ \chi_\lambda(n) := \frac{1-\lambda(n)}{2} = \begin{cases} 1 & \text{if } \ \lambda(n) = -1 \\ 0 & \text{if } \ \lambda(n) = 1 \end{cases}, \tag{3.3}$$

and study the average of the function $a_n s(n)$, namely,

$$S(N, \mathcal{H}, u) := \frac{1}{N}\sum_{n \sim N} a_n s(n), \ s(n) := s_{\mathcal{P}}(n) + \chi_\lambda(n+r), \ s_{\mathcal{P}}(n) := \sum_{i=1}^{k} \chi_{\mathcal{P}}(n+h_i). \tag{3.4}$$

We will compare this quantity with the average of the weights a_n, that is, with

$$A(N, \mathcal{H}, u) := \frac{1}{N}\sum_{n \sim N} a_n. \tag{3.5}$$

Our goal will be to show

$$S(N, \mathcal{H}, u) > A(N, \mathcal{H}, u) \tag{3.6}$$

which clearly implies the existence of at least one $n \sim N$ with $s(n) > 1$ and thereby the existence of either

(i) two primes of the form $n + h_i$, $n + h_j$ $(i \neq j)$ or

(ii) one prime $n + h_i$ and $\lambda(n + r) = -1$ $(r \neq h_i)$.

In the proof of (3.6) we can make use of Propositions 1 and 2 of [GPY], which we quote now as Lemmas 1 and 2 in the special case $\mathcal{H}_1 = \mathcal{H}_2$, $h_k \leq C(k) = O(1)$ as k will be bounded in our case. Constants c, C, c_i, C_i will be absolute unless otherwise stated and can be different at different occurrences. The same is true for constants implied by the $\ll$ or O symbols. The symbol o refers to the case $N \to \infty$, but it might also depend on k. The letters p, p_i will denote always primes.

The crucial singular series is defined for $\mathcal{H} = \mathcal{H}_k$ by

$$\mathfrak{S}(\mathcal{H}) := \prod_p \left(1 - \frac{\nu_p}{p}\right)\left(1 - \frac{1}{p}\right)^{-k} \geq c_0(k, \mathcal{H}), \tag{3.7}$$

for admissible $\mathcal{H}$ where $\nu_p = \nu_p(\mathcal{H})$ denotes the number of residue classes occupied by $\mathcal{H} \bmod p$. As stated in Theorems 3–6, we will always assume that $\mathcal{H}$ is admissible, that is,

$$\nu_p < p \quad \text{for} \quad p \in \mathcal{P}, \quad \text{equivalently} \quad \mathfrak{S}(\mathcal{H}) \neq 0. \tag{3.8}$$

With these notations we state (cf. (2.14)–(2.15) of [GPY])

Lemma 1. *If $R \ll N^{1/2}(\log N)^{-8(k+1)}$ then*

$$\frac{1}{N}\sum_{n\sim N} \Lambda_R(n;\mathcal{H},l_1)\Lambda_R(n;\mathcal{H},l_2) = \big(\mathfrak{S}(\mathcal{H}) + o(1)\big)\binom{l_1+l_2}{l_1}\frac{(\log R)^{k+l_1+l_2}}{(k+l_1+l_2)!}. \tag{3.9}$$

Lemma 2. *If $A > 0$ arbitrary, $R \ll N^{\vartheta/2}(\log N)^{-C(A,k)}$ then for any $h \in \mathcal{H}$ we have*

$$\begin{aligned}\frac{1}{N}\sum_{n\sim N} &\Lambda_R(n;\mathcal{H},l_1)\Lambda_R(n;\mathcal{H},l_2)\chi_{\mathcal{P}}(n+h) = \\ &= \big(\mathfrak{S}(\mathcal{H}) + o(1)\big)\binom{l_1+l_2+2}{l_1+1}\frac{(\log R)^{k+l_1+l_2+1}(\log N)^{-1}}{(k+l_1+l_2+1)!}.\end{aligned} \tag{3.10}$$

These lemmas take care of the evaluation of the averages of a_n and $a_n\chi_{\mathcal{P}}(n+h)$, so we have to deal only with the average of $a_n\lambda(n+r)$. This is in principle similar to the case $a_n\chi_{\mathcal{P}}(n+h)$ but in fact it is much easier. Although its evaluation reflects a deep property of the λ-function, it is (unlike the case of primes) a simple consequence of the analogue of the Bombieri–Vinogradov theorem, that is, of (2.2). In case $\vartheta > 1/2$ this is an unproved condition, which we assume in Theorems 5–8, while for $\vartheta = 1/2$ we can state it as

Lemma 3. *Relation* (2.2) *is true for $\vartheta = 1/2$.*

Its proof runs completely analogously to Theorem 4 of Vaughan [Vau] which is the analogous assertion for the Möbius function $\mu(n)$ in place of Liouville's function $\lambda(n)$. This implies easily the "analogue" of Lemma 2 for the λ-function.

Lemma 4. *If $A > 0$ arbitrary, $R \ll N^{\vartheta/2}(\log N)^{-C(A,k)}$ then for any $r \leq N$ we have*

$$\frac{1}{N}\sum_{n\sim N} \Lambda_R(n;\mathcal{H},l_1)\Lambda_R(n;\mathcal{H},l_2)\lambda(n+r) \ll_{k,l,A} (\log N)^{-A}. \tag{3.11}$$

Proof. For any squarefree m we have for the number $\nu_m = \nu_m(\mathcal{H})$ of solution of

$$P_{\mathcal{H}}(n) \equiv 0 \pmod{m} \tag{3.12}$$

by the Chinese remainder theorem

$$\nu_m = \prod_{p \mid m} \nu_p \le k^{\omega(m)} = d_k(m), \tag{3.13}$$

where $\omega(m)$ denotes the number of different prime divisors of m, $d_k(m)$ the number of ways to write m as a product of k integers. Therefore, interchanging the summation in (3.11) we obtain from (2.2) for the left-hand side of (3.11)

$$\begin{aligned} &\frac{1}{N} \sum_{d \le R} \sum_{e \le R} \frac{\mu(d)\mu(e) \left(\log \frac{R}{d}\right)^{k+l_1} \left(\log \frac{R}{e}\right)^{k+l_2}}{(k+l_1)!(k+l_2)!} \sum_{\substack{n \sim N \\ [d,e] \mid P_{\mathcal{H}}(n)}} \lambda(n+r) \ll \\ &\ll \frac{(\log R)^{2k+2}}{N} \sum_{q \le R^2} \Bigg(\sum_{q=[d,e]} 1 \Bigg) \nu_q \, E_{3N}(q) \end{aligned} \tag{3.14}$$

where (for $q \le M$)

$$E_M(q) := \max_{y \le M} \max_{a} \Bigg| \sum_{\substack{n \equiv a \pmod{q} \\ n \le y}} \lambda(n) \Bigg| \ll \frac{M}{q}. \tag{3.15}$$

Now, by (2.2) and (3.13), the sum over q in (3.14) is, similarly to (9.13) of [GPY],

$$\begin{aligned} &\ll \Bigg(\sum_{q \le R^2} \frac{d_{3k}(q)^2}{q} \sum_{q \le R^2} q_{3N} E_{3N}^2(q) \Bigg)^{1/2} \ll \\ &\ll \left((\log N)^{9k^2} N \cdot \frac{N}{\log^A N} \right)^{1/2} \ll N (\log N)^{(9k^2 - A)/2}. \end{aligned} \tag{3.16}$$

□

Using the notation

$$B := B(R, \mathcal{H}, k) := \frac{\mathfrak{S}(\mathcal{H}) \log^k R}{k!} \tag{3.17}$$

we have by Lemma 1

$$A(N,\mathcal{H},u) \sim B\left(1 + 2u + 2u^2 \cdot \frac{k+1}{k+2}\right). \tag{3.18}$$

On the other hand, Lemma 2 implies if $R = N^{(\vartheta-\varepsilon)/2}$

$$\begin{aligned} S_{\mathcal{P}}(N,\mathcal{H},u) := \frac{1}{N}\sum_{n\sim N} a_n s_{\mathcal{P}}(n) &\sim \\ &\sim \frac{Bk(\vartheta-\varepsilon)}{2}\left(\frac{2}{k+1} + \frac{6u}{k+2} + \frac{6u^2(k+1)}{(k+2)(k+3)}\right). \end{aligned} \tag{3.19}$$

Finally, Lemma 4 yields

$$S_\lambda(N,\mathcal{H},u) := \frac{1}{N}\sum a_n \chi_\lambda(n+r) \sim \frac{A(N,\mathcal{H},u)}{2}. \tag{3.20}$$

It follows from (3.3)–(3.4) and (3.17)–(3.20) that in order to show the crucial relation $S(N,\mathcal{H},u) > A(N,\mathcal{H},u)$ we have to find a value u such that

$$S_{\mathcal{P}}(N,\mathcal{H},u) > \frac{1+\varepsilon}{2}A(N,\mathcal{H},u) \tag{3.21}$$

which is satisfied if

$$k\vartheta\left(\frac{2}{k+1} + \frac{6u}{k+2} + \frac{6u^2(k+1)}{(k+2)(k+3)}\right) > 1 + 2u + \frac{2u^2(k+1)}{k+2}. \tag{3.22}$$

If we want to show Theorem 3 (in which case $\vartheta = 1/2$) with $k = 6$ we have to prove the existence of a u with

$$\frac{6}{7} + \frac{9u}{4} + \frac{7u^2}{4} - \left(1 + 2u + \frac{7u^2}{4}\right) > 0 \tag{3.23}$$

which is equivalent with $u > 4/7$. This proves Theorem 3.

In order to prove Theorem 5 we have to consider for $k = 2,3,4,5$ a quadratic inequality for u and calculate the value ϑ_0 for which the discriminant of the (in general really quadratic) formula equals 0. If $\vartheta_0 \geq 1$ the parameter k, the size of our set $\mathcal{H}$ is too small. If $\vartheta_0 < 1/2$, we have an unconditional solution. If $1/2 \leq \vartheta_0 < 1$, Theorem 5 is true for $\vartheta > \vartheta_0$.

Remark. Alternatively we can calculate the maximum of the ratio of the left- and right-hand side (taken without ϑ) of (3.22) and choosing the optimal value of u we get a lower bound for ϑ.

Remark. It is easy to see from (3.21)–(3.22) that working with the pure k-dimensional sieve corresponding to $l = 0$ in the weight function (3.1) (without using any other values l which corresponds to taking $u = 0$ in (3.2)) we are not able to prove any unconditional result, even for arbitrarily large value of k.

4 Proofs of Theorems 4 and 6

In case of the proofs of Theorems 4 and 6 we will choose one specific element $h_j \in \mathcal{H}$ and try to produce almost primes (with $\Omega(n + h_i) \leq C$) in all components $\{n + h_i\}_{i=1}^k$, and additionally either

(i) at least two primes $n + h_\nu$, $n + h_\mu$ with $\nu, \mu \in [1, k] \setminus \{j\}$, or

(ii) one prime $n + h_\nu$ with $\nu \in [1, k]$, $\nu \neq j$ and $\lambda(n + h_j) = -1$.

The starting point is to produce almost primes in each components, which was shown to be possible in [Pin] in such a way that using the weights (3.1)–(3.2), the total measure of those numbers n for which $\mathcal{P}_{\mathcal{H}}(n) = \prod_{i=1}^{k}(n+h_i)$ had a prime factor below N^η was negligible compared with the total measure of all $n \sim N$ if k and $\mathcal{H}_k$ were fixed, $N \to \infty$ and $\eta \to 0$. Denoting $P(m) = \prod_{p \leq m} p$ this was formulated (cf. Lemma 4 of [Pin]) in the following way.

Lemma 5. *Let $N^{C_0} < R \leq N^{1/(2+\eta)}(\log N)^{-C}$, $\eta > 0$. Then*

$$\sum_{\substack{n \sim N \\ (P_{\mathcal{H}}(n), P(R^\eta)) > 1}} \Lambda_R(n; \mathcal{H}, l)^2 \ll \eta \sum_{n \sim N} \Lambda_R(n; \mathcal{H}, l)^2, \tag{4.1}$$

where the constants C and the one implied by the $\ll$ symbol may depend on k, l and $\mathcal{H}$.

As remarked in [Pin] after the formulation of Lemma 4, the analogous quantity with the product of $\Lambda_R(n; \mathcal{H}, l_1)$ and $\Lambda_R(n; \mathcal{H}, l_2)$ with different values of l_1 and l_2 can be estimated by Cauchy's inequality and so we arrive at the more general relation

$$\frac{1}{N} \sum_{\substack{n \sim N \\ (P_{\mathcal{H}}(n), P(R^\eta)) > 1}} a_n(\mathcal{H}, u) \ll \frac{\eta}{N} \sum_{n \sim N} a_n(\mathcal{H}, u), \tag{4.2}$$

or equivalently (cf. (3.2))

(4.3)

$$\frac{1}{N}\sum_{\substack{n\sim N\\(P_{\mathcal{H}}(n),P(R^\eta))=1}} a_n(\mathcal{H},u)=\left(\frac{1+O(\eta)}{N}\right)\sum_{n\sim N} a_n(\mathcal{H},u)=:(1+O(\eta))A(N,\mathcal{H},u),$$

where the constants implied by the $\ll$ and O symbols may now depend on u, too. Since $s(n)\le k\le 9$, (4.2) immediately implies

(4.4)

$$S^*(N,\mathcal{H},u)=\frac{1}{N}\sum_{\substack{n\sim N\\(P_{\mathcal{H}}(n),P(R^\eta))=1}} a_n s(n)=S(N,\mathcal{H},u)+O(\eta A(N,\mathcal{H},u)).$$

This means that apart from a factor $1+O(\eta)$ the situation is the same as in the previous section, the only difference being that we have just $k-1$ components with primes instead of k. This means that choosing η as a sufficiently small constant in place of (3.22) we have now to ensure the inequality

$$(k-1)\vartheta\left(\frac{2}{k+1}+\frac{6u}{k+2}+\frac{6u^2(k+1)}{(k+2)(k+3)}\right)>1+2u+\frac{2u^2(k+1)}{(k+2)}. \tag{4.5}$$

In case of Theorem 4 we have $\vartheta=1/2$ and for $k=9$ we have to find a u with

$$\frac{4}{5}+\frac{24u}{11}+\frac{20u^2}{11}-\left(1+2u+\frac{20u^2}{11}\right)>0, \tag{4.6}$$

which is equivalent with $u>1.1$. This proves Theorem 4.

Theorem 6 can be shown similarly to Theorem 5, as described at the end of the previous section.

5 Concluding remarks

The more general formulation of Theorems 3–4 shows that apart from the small d's in Theorems 1–2 we obtain actually many different even values of d such that $\lambda(p+d)=1$ for infinitely many primes p. In fact, the lower density $\mathbf{d}(\mathcal{D}_0)$ of the corresponding set $\mathcal{D}_0$ of such d's is positive and it is easy to give an explicit lower bound for it as well. The argument is simple and similar to the one of Section 11 in [Pin]. In general, suppose k is bounded and $\mathcal{H}_k\subset[1,U]$, where we may assume $P:=P(k):=\prod_{p\le k}p\mid U$, since $U\to\infty$.

If we choose all elements from the set

$$\mathcal{M} := \{m \le U;\ (m,P)=1\}, \quad \text{where } M := |\mathcal{M}| = \frac{\varphi(P)}{p}U, \tag{5.1}$$

then $\mathcal{H}_k$ will be admissible since the zero residue class is not covered by $\mathcal{H}$ modulo any $p \le k$. Taking all choices for $\mathcal{H}_k$ we obtain $\binom{M}{k}$ even values of d, counted with multiplicity. A fixed difference d implies at most $M-1$ choices for the pair h_i, h_j with $h_i - h_j = d$ and afterwards we have $\binom{M-2}{k-2}$ choices for the remaining $k-2$ elements. This implies for the multiplicity of any d the upper bound $(M-1)\binom{M-2}{k-2}$, hence we obtain at least

$$\frac{M}{k(k-1)} = \frac{\varphi(P)}{P} \cdot \frac{U}{k(k-1)} \tag{5.2}$$

suitable even d's, hence with the notation (let $c = c(k)$ be a small fixed constant)

$$\begin{aligned} \mathcal{D}_0 &= \big\{2 \mid d, \#\{p; \lambda(p+d) = -1\} = \infty\big\} \\ \mathcal{D}_1 &= \big\{2 \mid d, \#\{p; \lambda(p+d) = -1,\ P^-(p+d) > p^c\} = \infty\big\} \end{aligned} \tag{5.3}$$

in the unconditional case $k=9$ this yields with the choice $j=k$ in Theorem 4

$$\mathbf{d}(\mathcal{D}_1) \ge \frac{1}{315}, \tag{5.4}$$

so that more than 0.6% of the even numbers have this property.

If we consider the conditional cases too, then the case $\vartheta > 0.927$ can be treated in a more efficient way, if we take into account that for any integers s and t with $0 < r < 3t$ the system $\mathcal{H}_3 = \{0, 2s, 6t\}$ is admissible, since $\nu_2(\mathcal{H}) = 1$, $\nu_3(\mathcal{H}) \le 2$, $\nu_p(\mathcal{H}) \le 3$ for $p > 3$. Let us choose $h_j = 6t$ in Theorem 6. Then at least one of the three numbers $2s$, $6t-2s$, $6t \in \mathcal{D}_1$. Let U be a large number and let

$$t^* = \max\{t;\ 6t \le U,\ 6t \notin \mathcal{D}_1\}. \tag{5.5}$$

Then clearly for every s at least one of $2s$ and $6t^* - 2s$ belongs to $\mathcal{D}_1$; further $6t \in \mathcal{D}_2$ if $t^* < t \le U/6$. This yields in total at least

$$\lfloor U/6 \rfloor - t^* + \left\lfloor \frac{6t^*}{4} \right\rfloor \ge \lfloor U/6 \rfloor \tag{5.6}$$

elements, hence

$$\mathbf{d}(\mathcal{D}_1) \ge 1/6 \ \text{ if } \ \vartheta > 0.927, \tag{5.7}$$

that is, essentially at least one third of all even integers belong to $\mathcal{D}_1$.

For $4 \le k \le 8$ choosing again $j = k$ in Theorem 6 we obtain from (5.2)

$$\begin{aligned}
&\mathbf{d}(\mathcal{D}_1) \ge 1/36 \quad \text{for} \quad \vartheta > 0.739,\\
&\mathbf{d}(\mathcal{D}_1) \ge 1/75 \quad \text{for} \quad \vartheta > 0.643,\\
&\mathbf{d}(\mathcal{D}_1) \ge 2/225 \quad \text{for} \quad \vartheta > 0.584,\\
&\mathbf{d}(\mathcal{D}_1) \ge 4/735 \quad \text{for} \quad \vartheta > 0.547,\\
&\mathbf{d}(\mathcal{D}_1) \ge 1/245 \quad \text{for} \quad \vartheta > 0.516.
\end{aligned} \tag{5.8}$$

Concerning the lower density of $\mathcal{D}_0$ in the case of Theorem 3, we start with an admissible $\mathcal{H}_k \subseteq [1, U]$ and with any even $r \in [1, U] \setminus \mathcal{H}_k$. This system yields at least one $d \in \mathcal{D}_0$ with either

(i) $d = h_j - h_i, \quad 1 \le i < j \le k$ or

(ii) $d = |r - h_i|, \quad 1 \le i \le k$.

We again start with the assumption $P(k)|U$, use notation (5.1), choose $\mathcal{H}_k \subset \mathcal{M}$ and obtain

$$\binom{M}{k}\left(\left\lfloor \frac{U}{2} \right\rfloor - k\right) \sim \frac{U}{2}\binom{M}{k} := Y \tag{5.9}$$

such pairs $(\mathcal{H}_k, r)$.

Any given d might arise in the way (i) at most $\frac{U}{2}(M-1)\binom{M-2}{k-2}$ ways as in (5.2). On the other hand if d arises as in (ii), then we have $\binom{M}{k}$ choices for $\mathcal{H}_k$, k choices for i and afterwards two choices for r, altogether $2k\binom{M}{k}$ possibilities. This gives for the multiplicity of d the upper bound (for both cases (i) and (ii) together)

$$\frac{U}{2}(M-1)\binom{M-2}{k-2} + 2k\binom{M}{k} = Y\left(\frac{k(k-1)}{M} + \frac{4k}{U}\right) := Z. \tag{5.10}$$

Hence the total number of different values $d \in \mathcal{D}_0$ is $\ge Y/Z$, and consequently

$$\mathbf{d}(\mathcal{D}_0) \ge \frac{Y}{Z} \cdot \frac{1}{U} = \left(\frac{ZU}{Y}\right)^{-1} = \left(4k + \frac{k(k-1)P}{\varphi(P)}\right)^{-1}. \tag{5.11}$$

This yields for the unconditional case $k = 6$ the estimate

$$\mathbf{d}(\mathcal{D}_0) \ge 2/273, \tag{5.12}$$

which means that more than 1.4% of all even numbers d belong to $\mathcal{D}_0$, equivalently satisfy $\lambda(p+d) = -1$ for infinitely many primes p. The conditional cases $k \le 5$ follow from (5.11):

$$\begin{aligned}\mathbf{d}(\mathcal{D}_0) &\ge 1/12 \quad \text{for} \quad \vartheta > 0.729,\\ \mathbf{d}(\mathcal{D}_0) &\ge 1/30 \quad \text{for} \quad \vartheta > 0.616,\\ \mathbf{d}(\mathcal{D}_0) &\ge 1/52 \quad \text{for} \quad \vartheta > 0.554,\\ \mathbf{d}(\mathcal{D}_0) &\ge 1/95 \quad \text{for} \quad \vartheta > 0.515.\end{aligned} \tag{5.13}$$

This means that under the strongest condition $\vartheta > 0.729$, for example, one sixth of all even integers d have the property that $\lambda(p+d) = -1$ for infinitely many primes p. On the other hand Corollary 1 shows under the same assumption $\vartheta > 0.729$ that for any even d either d or $2d$, consequently at least half of the even integers appear as the difference of two numbers among with at least one is a prime and the other has an odd number of prime factors.

A further interesting problem is that if we already know that $d \in \mathcal{D}_0$ or $d \in \mathcal{D}_1$ then what sort of lower bound can be given for the number of primes below N with the property $\lambda(p+d) = -1$. In case of $\mathcal{D}_0$ (Theorem 3) the proof implies only the weak lower bound

$$N \exp\left(-C \frac{\log N}{\log\log N}\right). \tag{5.14}$$

We can say much more in case of $\mathcal{D}_1$ (Theorem 4, or in the conditional case Theorem 6). If we have, namely, for a $k \in [3,9]$ with a given $n \sim N$ $P^-\big(P_{\mathcal{H}_k}(n)\big) > N^c$, then $P_{\mathcal{H}_k}(n)$ has $O_{k,c}(1)$ divisors, consequently

$$a_n(\mathcal{H}; u) \ll_u (\log R)^{2k} \ll (\log R)^{2k}, \tag{5.15}$$

since u was chosen bounded always.

Taking into account that we obtained in (4.4) finally (cf. (3.14)–(3.23))

$$\frac{1}{N} \sum_{\substack{n\sim N\\ (P_{\mathcal{H}}(n), P(R^\eta))=1}} a_n \Bigg(\sum_{\substack{i=1\\ i\neq j}}^{k} \chi_{\mathcal{P}}(n+h_i) + \chi_\lambda(n+h_j) - 1\Bigg) \gg_{k,u} B = \frac{\mathfrak{S}(\mathcal{H}) \log^k R}{k!}, \tag{5.16}$$

this implies by (5.15) that we found

$$\gg_{k,u} \frac{\mathfrak{S}(\mathcal{H}) N}{\log^k N} \tag{5.17}$$

integers $n \in [N, 2N)$ with the property

$$\sum_{\substack{i=1\\ i\neq j}}^{k} \chi_{\mathcal{P}}(n+h_i) + \chi_\lambda(n+h_j) > 1, \quad P^-(P_{\mathcal{H}}(n)) > R^\eta \geq N^c \tag{5.18}$$

since $R \geq N^{1/5}$ was chosen. We can take here unconditionally $k = 9$, while for $\vartheta > 1/2$ we obtain the lower estimates

$$\frac{\mathfrak{S}(\mathcal{H})N}{(\log N)^{C_2(\vartheta_0)}} \quad \text{if} \quad \vartheta > \vartheta_0, \tag{5.19}$$

where the values for $C_2(\vartheta_0)$ are given in the formulation of Theorem 6. We remark that we can omit the dependence on k, u and $\mathcal{H}_k$ since $k \leq 9$, u is bounded and for any admissible $\mathcal{H}_k$ we have

$$\mathfrak{S}(\mathcal{H}_k) \geq \prod_{p\leq 2k} \frac{1}{p}\left(1-\frac{1}{p}\right)^{-k} \prod_{p>2k}\left(1-\frac{k}{p}\right)\left(1-\frac{1}{p}\right)^{-k} \gg_k 1. \tag{5.20}$$

On the other hand Selberg's sieve (cf. Theorem 5.1 of [HR] or Theorem 2 in §2.2.2 of [Gre]) yields for the number of n's in $[N, 2N)$ with $P^-\big(P_{\mathcal{H}_k}(n)\big) > N^c$ the upper bound

$$C(k)\frac{\mathfrak{S}(\mathcal{H}_k)N}{(\log(N^c))^k} = C'(k)\frac{\mathfrak{S}(\mathcal{H}_k)N}{(\log N)^k}, \tag{5.21}$$

since c can be chosen as a small constant depending on k. This shows that the lower estimate (5.17) is sharp up to a constant factor and also that a positive proportion of all almost prime k tuples, that is all $n \in [N, 2N)$ with

$$P^-(n+h_i) > N^c \quad \text{for} \quad 1 \leq i \leq k \tag{5.22}$$

satisfy that we can find among the elements $n + h_i$ at least one prime and another one with an odd number of prime factors ($\lambda(n + h_j) = -1$).

We can summarize the unconditional case (cf. (5.4) and (5.17)) as

Theorem 9. *We have an infinite set $\mathcal{D}_1$ of positive even numbers d with lower density $\geq 1/315$, including at least one $d \leq 30$, absolute constants C, c, c' and an* odd *integer b, such that there are at least*

$$c'\frac{N}{\log^9 N} \tag{5.23}$$

primes up to N ($N > N_0(d)$) with $\Omega(p + d) = b$, $P^-(p + d) > p^c$.

The fact that for a positive proportion of the almost prime k-tuples with (5.22) we have at least one prime among $n + h_i$ and another one with $\lambda(n + h_j) = -1$, make possible a proof of the following extension of Theorem 9.

Theorem 10. *We have absolute constants C, c, an* odd *$b \le C$ and a set $\mathcal{D}_1$ of even numbers with lower density $\ge 1/315$, including at least one $d \in \mathcal{D}_1$ with $d \le 30$ such that the set $\mathcal{P}(d)$ of primes p with $\Omega(p + d) = b$, $P^-(p + d) > p^c$ contains arbitrarily long arithmetic progressions.*

We will omit the proof of Theorem 10, since it follows the same line of arguments as [Pin], where it was proved that if the primes have a distribution level $\vartheta > 1/2$ then we have a positive $d \le C(\vartheta)$ such that there exist arbitrarily long arithmetic progressions of primes p such that $p + d$ is also a prime (in fact the one, following p) for all elements of the progression.

References

[Bom] E. Bombieri, The asymptotic sieve, *Rend. Accad. Naz.* XL (5) 1/2 (1975/76), 243–269.

[Che1] Jing Run Chen, On the representation of a large even integer as the sum of a prime and the product of at most two primes, *Kexue Tongbao* **17** (1966), 385–386 (Chinese).

[Che2] Jing Run Chen, On the representation of a large even integer as the sum of a prime and the product of at most two primes, *Sci. Sinica* **16** (1973), 157–176.

[GPY] D. A. Goldston, J. Pintz, C. Y. Yıldırım, Primes in tuples I, *Ann. of Math. (2)* **170** (2009), no. 2, 819–862.

[Gre] G. Greaves, *Sieves in Number Theory*, Springer, 2001.

[HR] H. Halberstam, H.-E. Richert, *Sieve Methods*, Academic Press, London, 1974.

[Hil] A. J. Hildebrand, *Erdős' problems on consecutive integers, Paul Erdős and his Mathematics I*, Bolyai Society Mathematical Studies **11**, 305–317, Budapest, 2002.

[Iwa] H. Iwaniec, Prime numbers and L-functions, *International Congress of Mathematicians*, Vol. I, 279–306, Eur. Math. Soc., Zürich, 2007.

[Pin] J. Pintz, Are there arbitrarily long arithmetic progressions in the sequence of twin primes?, preprint, arXiv:1002.2899.

János Pintz
Rényi Mathematical Institute of the Hungarian Academy of Sciences
Budapest
Reáltanoda u. 13–15
H-1053 Hungary
E-mail: pintz@renyi.hu